DOCUMENT RESUME

ED 217 098 UD 022 237

AUTHOR Madhere, Serge
TITLE Issues in Remedial Program Evaluation. ORET Report No. 3.
INSTITUTION Newark Board of Education, NJ. Office of Research, Evaluation and Testing.
PUB DATE Oct 81
NOTE 60p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Achievement Gains; Comparative Analysis; *Compensatory Education; Elementary Education; Individualized Instruction; Learning Laboratories; Mainstreaming; *Program Effectiveness; *Teaching Methods
IDENTIFIERS *Elementary Secondary Education Act Title I; *Newark School System NJ

ABSTRACT

This report consists of the descriptions and results of two studies carried out to examine Title I programs in the Newark, New Jersey, public schools. The Sustaining Effects Study focused on two groups of pupils: those who participated in Title I programs for two years (1979 to 1982), and those who received remediation for one year (1979-80) and because of their success did not need assistance in 1980-81. An extended Model-C analysis was performed for the first group and an Extended Model-A analysis was performed for the second group. Results show sustaining effects in grade 6 for both reading and math, in grade 5 for math only, and in grade 4 for neither component. The Instructional Strategy Study aimed at determining which instructional strategy (regular classroom, pull-out, or lab) has been most effective with Title I pupils. Twelve classes from grades two to seven were considered at four different schools. The impact of student overall academic standing and general program status was controlled. Results of the analysis of variance performed showed that the pull-out method tends to be the most effective, although where Title I enrollment was small, the lab approach was preferable. (Author/GC)

ORET REPORT NO. 3

ISSUES IN REMEDIAL PROGRAM EVALUATION

Prepared by

Office of Research, Evaluation and Testing
Program Monitoring and Evaluation
Newark Board of Education
2 Cedar Street
Newark, New Jersey 07102

OCTOBER 1981

ORE REPORT NUMBER 3

ISSUES IN REMEDIAL PROGRAM EVALUATION

Serge Madhere, Ph.D.

Newark Board of Education
Division of Program Development
Office of Research and Evaluation
2 Cedar Street
Newark, New Jersey 07102

OCTOBER 1981

TABLE OF CONTENTS

LIST OF TABLES

TABLE OF CONTENTS (Cont'd.)

Acknowledgements

I would like to express my gratitude to a few people for their assistance in completing this task. My appreciation goes first to my supervisor, Dr. Ram Durga, for his continuing encouragement and support. In the early stage of the study, a lengthy discussion with him and Dr. Bruce Hartman helped me very much clarify my own thinking about sustaining effects. Dr. Hartman even consented to do the preliminary computer runs for the Model-A analysis.

Thank to Tracey Chamberlin who perfected the grid for Title I project setting description (in Appendix B), and saw through it that the needed information reach the Office of Research and Evaluation. Here I must mention, of course, the Title I project coordinators for most of the data gathering on these and other research projects. Thanks also to Dr. Hope Hartman-Haas for her patience in editing draft after draft of this paper. Last but not least, Jacquelyn Henry and Pegggy Williams deserve by gratitude for the fine job they did, one in typing, the other in desigining the cover for this booklet. It is certainly nice to be part of such a team.

FOREWORD

This monograph was first conceived as an extension of the 1981 Title I evaluation report. The original report, as required, documented student progress, (or lack of it), at the various grade levels, and for the different programs coming under the umbrella of the Title I project. But a need was felt to go beyond that kind of summative evaluation, beyond the issue of whether something was achieved, to address the question of how and how well it was achieved. These are simple questions, but the answers to them are far from being simple. Indeed, they require a shift of focus from educational objectives to educational practices; and the latter are always more complex than the former.

The primary objective of the Title I program is to improve student achievement in basic skills, within a year, beyond the initial, unsatisfactory level of performance. But the educational practice tied to that objective can be shaped according to any one of three different instructional strategies; and it is hardly known which one is the most effective. A secondary objective of the Title I project is to make the achievement gains consistent and permanent. But the educational practice often entails the selection of pupils in and out of the remedial program, after one year of service; this amounts to the administration of a double dose of remediation to some students, a single dose to others; and it is hardly known where exactly the pay-off is.

That kind of practical concern first defined the focus and the limits of the investigation. But, even with those limits set, it soon became evident that the two educational practices under consideration — instructional strategy and continuity/discontinuity of services --branched out into a number of other areas such as ability grouping, teacher/pupil interaction, curriculum coordination. So, to understand the educational practices, one had to keep the latter variables in full view. However, since these variables were not systematically built into the Title I project design, an examination of them could not be done within an evaluation framework. In other words, the task changed from one of pure evaluation to one of (exploratory) research.

Taking a research orientation required some conceptual and methodological elaborations. Neither might seem immediately relevant to the Title I program. But both are needed when one is to deal with such unsettled issues as the long-term impact of, or the instructional strategies for compensatory education. Both are needed if one is to reconcile the results of research and evaluation with educational principles. However, this monograph goes beyond a simple theoretical statement. Practical and immediate applications of the findings are offered. This does not mean that the last word is said on the issues. Much work is needed on these and other meaningful questions concerning the Title I program. But, at this point, who would discourage any effort to make educational practices become more educational, and look less like "practice runs"?

ABSTRACT

1. The Sustaining Effects Study

This study focuses on two groups of pupils: those who have been Title I participants for two consecutive years (1979-1981); and those who have received remediation for one year (in 1979-80), and because of their success, did not need assistance in 1980-81. An extended Model-C analysis was performed for the first group; an extended Model-A analysis was performed for the second group. The results show sustaining effects in grade 6 for both reading and math; in grade 5 for math only; and in grade 4 for neither component.

2. The Instructional Strategy Study

This study aims at determining which instructional strategy (regular classroom, pull-out or lab) has been most effective with Title I pupils. Twelve classes from grades 2 to 7, were considered at four different schools. The impact of such variables as student overall academic standing and general program status was controlled. An analysis of variance was performed, which tested also for interactions among the variables. The results show that the pull-out method tends to be the most effective. However, where the Title I enrollment is relatively small, the lab approach is to be preferred. As a practical application of these findings, a decision-tree for setting up the remedial reading program has been developed.

Instructional Strategy and Student Progress in Reading

1. Which instructional strategy -- regular classroom, pull-out, or lab -- leads to the greatest rate of progress in reading by Title I pupils?
2. Should the same approach be used with pupils in need of remediation in both reading and math, and with those deficient in only one of those basic skills?
3. Is the degree of effectiveness of each instructional strategy uniform in any school setting, or does it depend on the remedial program status?
4. If more than one instructional approach has to be followed, which combination should be adopted?

Perspective

In setting up a remedial instruction program, such as Title I, school administrators usually have three options. Under the first option, the remedial instruction is offered in the regular classroom. Under the second option, commonly known as "pull-out", students in need of remediation are taken to a separate classroom for instruction possibly with a subject-matter specialist. The third option provides for the remedial learning experience in a lab, where at least seventy-five percent of the instruction is supported by a machine. Each option clearly implies a different instructional strategy. When all three are present within a district or within a school, the Title I project becomes a multi-faceted program. To fully understand the program, each facet must be studied.

It seems that the selection of a particular instructional strategy or format is rarely in response to concerns about student progress or the demands of curriculum construction. Some project administrators tend to prefer the regular classroom approach simply because it is the least expensive and requires no special scheduling or coordinating plan from classroom to classroom. On the other hand, "the practice of pulling Title I eligible pupils out of their regular classroom, as Glass and Smith found out, is an artifice created by schools to satisfy (funding sources) regulations concerning "supplementing not supplanting instruction" (Glass and Smith, 1977, p. 6). Concerns about the latter point seem to be a compelling factor. In the Newark school district, the format most often implemented at all grade levels is the pull-out (see Table A 1).

Questions regarding whether different instructional strategies result in different rates of progress in reading or math have been raised from time to time (Glass and Smith, 1977; NIE, 1976,1977). Many studies have contrasted the regular classroom and the pull-out approaches. The expectations have been that the regular classroom setting would be more conducive to student achievement. Such expectations have been based on two tenets of educational wisdom: a) In the regular classroom, opportunities exist

for informal peer-tutoring, the less advanced student receiving assistance from the more able, thus improving his/her performance. b) The pull-out strategy may be the moral equivalent of quarantine. It leads to separatism, creates a label, and with it comes lower expectations about achievement. The problem is compounded for the Title I students because "both labeling of a pupil as "Title I eligible" and emphasizing such a label by removing him from the regular classroom create strong expectancy biases in teachers and pupils alike" (Glass and Smith, p. 28). Achievement is thus in jeopardy.

The empirical evidence to support the above contentions is scant and sometimes contradictory. A NIE study (1977) found no significant difference between the regular strategy and the pull-out strategy in their impact on first-grade reading. Some counter-arguments have been developed in favor of the pull-out approach. It is pointed out that a) such a strategy lowers the pupil/teacher ratio, increases thus the amount of teacher-pupil interaction. And that, as everyone would agree, has a positive influence on student rate of progress. b) Instruction in the pull-out setting is often offered by a subject-matter specialist, someone with more advanced skills/training than the regular classroom teacher. And staff educational level is also a good correlate of pupil achievement. So, far from creating any kind of dislocation, the relocation/pull-out format allows for greater specificity and intensity of learning.

Which one of these two contrasting points of view is more valid? Before that can be answered, one may have to cast the problem in different terms, i.e., address the more basic question underlying the debate. That question is: Of two sources of influence -- that of teachers or that of peers -- which one is the more critical in determining the progress of under-achieving students? Proponents of the first point of view seemingly emphasize the role of the peer group; proponents of the second point of view are insisting on the role of the teacher.

It is the opinion of this author that there is no unique answer to the question. Experience, logic, and theory suggest that it is more fruitful to look at the problem in

light of some other variables such as student academic ability or overall achievement, and the global program context.

Student overall achievement. -- It is possible that the under-achieving pupil with the greatest needs can benefit only through the intense interaction with the teacher, i.e., in the pull-out setting, while the pupil whose academic performance is not consistently low can function/advance in the regular classroom with some assistance from the peers. How does one make the distinction between the two subgroups of under-achieving pupils? If a child has to be given remedial instruction in both reading and math, then his/her overall academic performance is deemed low; he/she is in greater difficulty than another child who needs remediation in reading only. Children in the latter situation can be referred to as the single-need pupils, while children in the former situation can be designated at the multiple-need pupils. What is being suggested here is that a given instructional approach takes greater or lesser significance in context. And, true to the notion of heterogeneity in the Title I population, the point of interest becomes the possibility of an interaction between the variables instructional format and pupil overall achievement.

Program context. -- Variations in overall performance level define what one may call the classroom academic context. Beyond that immediate context, one must look at the global school context which may be more directly related to the issue of labeling. The argument goes that the pull-out strategy, because it tends to reinforce the difference between remedial program participants and non-participants, stigmatizes the first group, and that has a negative impact on achievement. But, in general, stigmatization is not an automatic consequence of differentiation and even labeling. As one knows from the theory of deviance (Festinger, 1954; Merton, 1957), within a social structure, stigmatization is usually aimed at a subgroup with a marginal/minority status. Applying this principle to the school situation, one would say that labeling may be equivalent to stigmatization mainly in an environment where the Title I students

represent a relatively small group. But where they are a majority or quasi-majority, they may become an important social-reference group; and, in that context, the sting of labeling may just dull out. Since, from building to building, there is great variation in the percentage of the student population enrolled in compensatory education, this factor of school context or program status takes some significance. The regular classroom arrangement may be more conducive than the pull-out where the Title I program has a marginal status, and labeling/stigmatization is stronger; it may not make a difference where the remedial program participants are a majority or quasi-majority, i.e., the program is central to the school structure. Once again, what is being suggested is the possibility of an interaction between the variables instructional format and remedial program status.

All in all, the present study will try to answer 3 questions:

1. Under what instructional arrangement -- regular classroom, pull-out, or lab -- is the greatest rate of progress achieved by Title I students?
2. Are single-need pupils and multiple-need pupils sensitive to different instructional strategies?
3. Is the degree of effectiveness of each instructional strategy uniform in any school setting, or is it a function of the program status?

Method

Sample

Data on instructional format was collected throughout the Newark school district in February 1981. At that time, a grid was constructed and circulated, asking the principals and the Title I program coordinators to document the type(s) of instructional arrangement adopted at each grade level in their school. The three common alternatives were presented: regular classroom, pull-out, and lab. Within any grade, if more than one Title I subgroup was formed, and/or more than one approach was used, that also had to be indicated. For the present study, four public elementary schools were selected. At two of these schools, the number of Title I students represents less than 36% of the school population; the status of the Title I program, in this context, can be considered as marginal. At the other two schools, Title I students account for more than 45% of the total enrollment; the Title I program, in that context, can be regarded as central to the school structure. At each school, three classes were selected to represent the primary grades (2-3), the intermediate grades (4-5) and the advanced grades (6-7). At each educational level, two subgroups were considered: one in which the students are deficient in both reading and math (multiple-need group); another which include students deficient in reading only (single-need group). One hundred and seventy nine pupils are part of the first subgroup; the remaining seventy belong to the second group. So, the total sample includes a total of 249 pupils in 12 classes, and 24 subgroups.

To measure the dependent variable, rate of progress of Title I students, both the 1980 and 1981 reading scores on the Metropolitan Achievement test were examined. These scores were all transformed into normal curve equivalents (NCE's), making them comparable from grade to grade. For each pupil, the 1980 NCE was subtracted from the 1981 NCE, yielding an index of change/progress in reading over the year. At each school and for each grade considered, these change indexes were averaged to give the

TABLE A1

Number of Schools Using Each Educational Setting

Language Experience Program

Grade Level	Regular Classroom		Pull-out		Laboratory	
	Public	Nonpublic	Public	Nonpublic	Public	Nonpublic
1	36	10	11	3	0	0
2	15	3	27	4	8	1
3	10	5	32	6	12	2
4	6	3	27	5	15	2
5	7	3	20	5	13	2
6	8	3	23	7	9	2
7	6	1	16	6	8	0
8	6	1	9	4	3	0
9	0	0	0	0	7	0

Data reported in February 1981 by Title I Project Coordinators

mean rate of progress. The calculations bore on the averages of progress rate, so the class subgroup served as the unit of analysis.

Design

A split-plot factorial design (Kirk, 1968) was adopted to organize the data, and control for as many sources of variance as possible. The complete lay-out is presented in Table A2. As one can see, the variables program status, instructional format, and degree of academic deficiency are completely crossed, i.e., every possible combination of one of these variables with another appears in the design. Possible differences between grade levels are kept in check, since the various grade levels (primary, intermediate, advanced) are represented under each instructional format, each program status category, and each school. However, variations from school to school are only partially controlled. Indeed, each school serves as a block and provides observations under each instructional option; but because of the nature of the variable program status -- at any given school, the Title I program can have only one status: marginal or central -- school blocks could only be nested under the variable program status. It is because of the need to resort to nesting that the split-plot design was preferred over a ramdomized-block factorial approach.

Analysis

A three-way analysis of variance procedure was adopted. The general model was a fixed-effects model with replication, which means that no generalizations were intended beyond the variable categories specifically included in the design. The hypotheses were tested that there were no main effects from any of the variables, but that the two-way interactions (involving pairs of variables) were significant. Concretely, all this means that:

Main Effects: 1 - Considered by itself, program status does not create any significant difference in pupil achievement.

2 - Considered by itself, no instructional strategy yields greater results than any other.

3 - There is no significant difference in reading rate of progress between the single-need pupils and the multiple-need pupils.

Interactions:

4 - The effectiveness of a given instructional strategy varies with the status of the Title I program.

5 - The effectiveness of a given instructional strategy varies with the degree of academic deficiency shown by students.

Results

The main results of the analysis are reported in Table A3. For each of the 24 subgroups, the rate of progress is given in Table A2. Thirteen of the subgroups have improved their performance from one year to the next, while eleven have performed below the level of expectation. A calculated multiple-eta square of .718 indicates that approximately 72% of the variance in Title I student progress in reading can be attributed to the joint effect of program status, instructional format, and pupil degree of academic deficiency. But when one tries to break down this value into its various components, it becomes obvious that not much of the impact is attributable directly, as main effects to the three variables. Table A3, presenting the summary of the analysis of variance, shows that:

1. The F-statistic, for the variables program status and degree of academic deficiency, is equal to 4.8 and .62 respectively, Neither value is significant at the .05 level. So, the first and third hypotheses are confirmed: considered independently, program status or academic deficiency in another subject-matter does not impact significantly the rate of progress in reading.
2. The second hypothesis, however, has to be rejected. The F-value associated with the variable instructional format is equal to 5.20, which is significant at the .05 level. Based on the value of omega-square calculated for that variable, it can be concluded that the selection of instructional approach reduces by approximately 11% the "uncertainty" about student rate of progress. When the overall means obtained under each instructional option are compared (see margin of Table A4), one notices that the pull-out approach is the only one yielding a positive figure (4.92). So, overall, the latter strategy is the most effective one.
3. Of the two interaction hypotheses formulated, confirming evidence is obtained for only one. First of all, none of the instructional approaches seems more effective with one caliber of students, and less appropriate for another caliber of students.

PROGRAM STATUS	SCHOOL/GRADE	INSTR. FORMAT	ACADEMIC NEED	
			MULTIPLE	SINGLE
MARGINAL	A/2	Regular	-7.7	-11.1
	B/5	Classroom	-1.1	-16.7
	A/4	Pull-out	1.4	-1.7
	B/3		-1.9	5.8
	A/6	Lab	2.7	6.0
	B/7		11.8	-9.9
CENTRAL	C/2	Regular	-3.2	9.6
	D/7	Classroom	6.1	6.0
	C/7	Pull-out	13.7	8.4
	D/3		9.3	4.4
	C/5	Lab	-4.9	3.2
	D/4		-8.3	-13.7

TABLE A2 - Progress in reading of Title I pupils in 3 different instructional settings at 4 different schools (Split-plot Factorial Design)

PROGRAM STATUS	INSTRUCTIONAL FORMAT		
	REGULAR CLASS	PULL-OUT	LAB
Marginal	-9.15	.9	2.65
Central	4.6	8.95	-5.9

TABLE A4 - Rate of progress as a function of instructional format and program status

SOURCE	SUM OF SQUARES	DEGREE OF FREEDOM	MEAN SQUARES	F VALUE	SIGNIF.
Between Groups	1055.95	11	---	---	---
1. Status	117.04	1	117.04	4.8	n.s.
2. Format	253.64	2	126.82	5.2	.05
3. Interaction (1)x(2)	539.13	2	269.57	11.07	.01
4. B.G. Error	146.13	6	24.35	---	---
Within Groups	557.03	12	---	---	---
5. Need	31.74	1	31.74	.62	n.s.
6. Interaction (1)x(5)	60.17	1	60.17	1.17	n.s.
7. Interaction (2)x(5)	7.96	2	3.98	.08	n.s.
8. Interaction (1)x(2)x(5)	148.36	2	74.18	1.44	n.s.
9. W.G. Error	308.80	6	51.46	---	---
Total	1612.98				

TABLE A3 - Analysis of variance data

The F-value for the interaction between these two variables is less than 1, so not significant. The pull-out approach holds the advantage at either level of academic deficiency. On the other hand, it is clear that the instructional strategies operate differently in one global school context or the other. The F-test for that interaction (hypothesis #4) yields a value of 11.07, which is significant beyond the .01 level. Corresponding to it is an omega-square value of .274.

The task is now one of finding out which school context offers the better match to each instructional strategy. To facilitate that kind of comparison, the results are condensed in Table A4. a) When one contrasts the regular classroom approach to the lab approach, one notices that the former is more effective in a school where Title I enrollment is substantial (4.6 vs -5.9), while the latter gives better results in a school where the Title I pupils are a minority (-9.15 vs 2.65). b) The same pattern of results is observed when the comparison bears on the pull-out and the lab approaches. In a school with a large Title I population (in proportion to the whole student body), the pull-out method leads to greater progress (8.95 vs -5.9), while the lab approach is again to be preferred when the size of the group in remediation is limited (.9 vs 2.65). c) When the regular classroom format is compared to the pull-out format, no reversal is observed in the achievement trend. Regardless of the program status within the global school context, the latter instructional strategy is superior.

Although no specific hypotheses were formulated concerning a three-way interaction (involving all three variables) or one between program status and student degree of deficiency, it is worth mentioning that neither component yields a significant F test.

Discussion

The preceding analysis clearly indicates that, overall, the most effective instructional strategy for improving the reading performance of Title I pupils is the pull-out. A NIE study (1977) had found similar results for achievement in mathematics at the third grade level. Glass and Smith (1977) had doubted these results. But the reality seems inescapable. The greater strength of the pull-out approach is due to the fact that it consistently yields positive outcomes, while the other two strategies are effective under certain conditions, ineffective in another context. The superiority of the pull-out format over the regular classroom format may reflect a more systematic effort in designing and following the remedial instruction sequence. In that sense, the approach represents more than just an attempt to placate federal or state monitors about "supplementing rather than supplanting" instruction. Another possible explanation for the strong impact of pull-out can be formulated in terms of general experience: since most school districts have, over the years, adopted predominantly the pull-out format, it is possible that instructors have developed greater expertise dealing with under-achieving pupils in that setting. The greater experience would lead to more consistency, and positive results.

Consistency is the term to be retained in the foregoing discussion. If one's purpose is to maximize effectiveness, then another dimension of the picture must be simultaneously considered :the Title I program status, or the extent to which it is marginal or central to the school structure. That is the real meaning of the strong interaction between instructional format and program status.

The pull-out appears most effective in an environment where the Title I pupils constitute a majority or quasi-majority. In such a context, the impact of labeling may be meaningless. So, pupils with remedial needs are getting the full benefits from a curriculum designed and/or conducted by a subject-matter specialist. That kind of professional support seems to weigh much more than a possible peer-tutoring, informal

at any rate, that occurs in the regular classroom. The latter approach, however, turns out to be more beneficial than the lab, in the present context. This finding, at first surprising, is in line with results obtained by S. Hartley (cited by Glass and Smith) at the University of Colorado. The latter author has observed that the impact of tutoring (which may happen in the regular classroom) is "greatly superior to the effects of much more expensive instructional methods such a computer-assisted instruction, individualized learning packets, and programmed instruction" (Glass and Smith, p. 28).

However, that eventual impact of peer-tutoring may be at work only when labeling is absent or minimized. Indeed, in an environment where the Title I pupils form a relatively small minority (less than one-third of the student body), the regular classroom approach becomes the least effective strategy and the lab the r. :t. This finding implies two things: a) despite the large number of advanced students afforded by such a context, the opportunities for peer-tutoring -- which would lead to improved performance by the remedial program participants -- do not materialize in the regular classroom. b) Since labeling and the concomittant stigmatization may be at their maximum in such a context, the best strategy, when removing pupils from their regular classroom is to take them to a lab. The pull-out approach and the lab present some common features, (both require the relocation of pupils), but the instructional activities may be quite different; and the apparent sophistication of the lab may contribute to minimizing the impact of labeling.

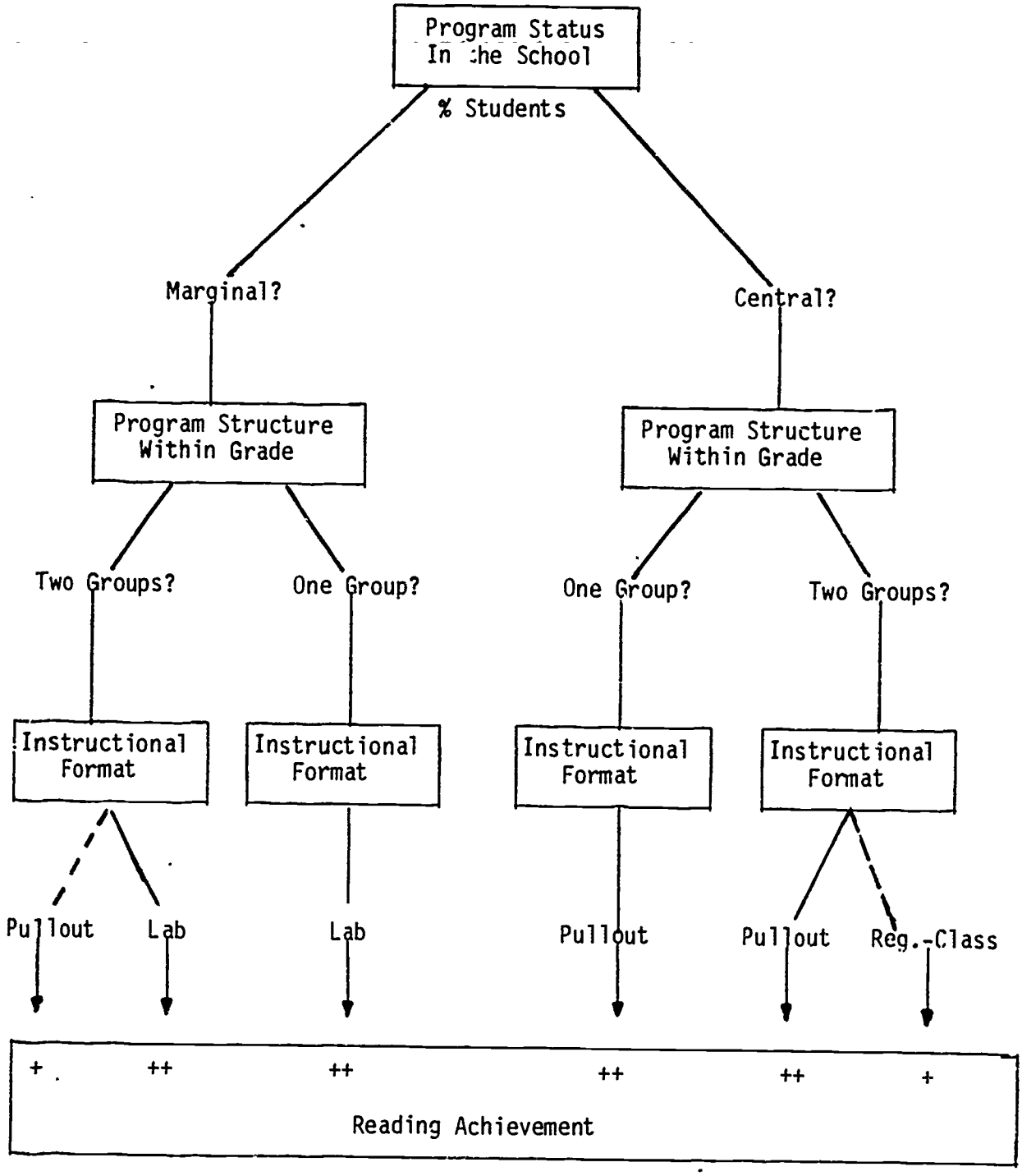
Recommendations

For practical purposes, the conclusions outlined above can be translated into a decision-tree for setting up a remedial reading program. This application takes into consideration the possible formation of more than one Title I group per grade level (due to a large number of eligible students in that grade), and the consequent need for the principal/program coordinator to follow more than one instructional strategy. The entire decision-tree is laid out on the following page. It shows that three different

Decision-Tree

On Instructional Strategy

For a Remedial Reading Program



features of a program (in boxes) have to be specified: the first one concerns the program status in the school, the second is the program structure within a given grade level, the third one pertains to the selection of a particular instructional format. The likely outcome of the selection is indicated in the rectangle at the bottom of the tree. Let's illustrate the use and usefulness of that decision-tree. By looking at the percentage of students in the school enrolled in the Title I program, the principal/project coordinator establishes whether the status of the program is marginal or central to the school structure. Once that determination is made, he/she examines the enrollment at each grade level to see if one or two Title I subgroups must be formed. If there is only one group, the decision is straightforward: lab (if the program has a marginal status), pull-out (if the program is central to the school structure). The degree of effectiveness is, in each case, indicated by a double or a single plus sign in the outcome box, reading achievement.

This paper deals mainly with the issue of instructional format. Much more information needs to be gathered on the precise instructional activities and materials adopted under each instructional format. That kind of information would make possible the preparation of a similar decision-tree for the instructor and not just for the administrator. In the short and the long run, the pupils in need of remedial instruction would benefit from such an endeavor.

Sustaining Effects of the Title I Program

1. How much of the gain, realized through remedial instruction, do successful Title I pupils maintain, after they leave the program?
2. Does it take two years of program participation, rather than one, for most Title I students, to show a gain in basic skills?
3. In any given year, should evaluation of student progress be based on their most recent performance only, or should it include achievement in more than one earlier grade?

Introduction

Program evaluation is often conceived and/or executed as a one-stage study which throws much light on the intervention period between pretesting or needs assessment and posttesting. But what happens after the posttesting date is anybody's guess. Even when the project is continued from year to year, the evaluation of each treatment period is done independently. The need for assessing the long-term impact of a program, though recognized by educators and policy makers, is not always addressed. Of course, there are a few longitudinal studies dealing mainly with the impact of early childhood education (Bronfenbrenner, 1974; Weinberg, 1979; Wolff and Stein, 1966). But, in general, the research effort has not been commensurate to the stated importance of the problem.

It is a truism to say that the evaluation of a project's long term impact or sustaining effects is necessarily a two-stage investigation. First, one must determine the most immediate effects at the conclusion of the intervention; that assessment is then repeated at a later point in time. What needs to be underlined, however, is that what takes place during the post-treatment period is as important as the treatment itself. In the regular school setting, one of two things is likely to occur: a) following the initial intervention, (and hopefully because of it), a number of students have made enough progress to be dispensed of the special instructional program, b) another group of students, less successful, have to be assigned to the program for another year.

Because of that treatment continuity/discontinuity factor, a study of a project's sustaining effects must make a distinction between the cumulative effects and the carry-over effects. When one is focusing on former participants, pupils who have successfully "graduated" from a program (treatment discontinuity) one is dealing with carry-over effects; the question of interest is then how much of the gain realized, as a

result of prior remedial instruction, is maintained in the next-higher grade. When one considers a group of pupils enrolled two consecutive years in a special program (treatment continuity), one has to pay attention to the cumulative effects of the project; the question of interest is whether a significant improvement in student achievement requires the longer program participation.

The purpose of this study is to evaluate both the cumulative and carry-over effects of the ESEA Title I program in the Newark school district, for students in grade 4, 5, and 6. To answer the two questions posed in the preceding paragraph, two different methodological approaches are necessary. In order to assess cumulative effects, an approach similar to the regression projection model (known as model -C1) but adopting a path analytic interpretation, has been developed and applied. In order to assess carry-over effects, a norm-referenced model (known as model - A1) has been used. A final judgement on the program's sustaining effects will integrate the results of both analyses.

Treatment Continuity: Cumulative Effects

It has repeatedly been observed that the majority of the students enrolled in a special program such as Title I do not make enough progress within a year to be dispensed of the supplemental instruction in the next higher grade. Sometimes, no significant progress can be detected at all. These students are likely to be automatically reassigned to the remedial program. But it is not always known whether there is any benefit to that double dose of remedial instruction. The question is: are the students better off after the second year? In other words, does it take two years rather than one for them to improve their reading or math achievement beyond the level that could be expected or brought about through the regular classroom instruction alone?

Design

To determine whether the Title I program had achieved that objective, a special model, based on a regression-projection design (Tallmadge and Horst, 1975), was developed. That methodological approach is a modification of the better known Evaluation Model C1. The latter is a 3-step procedure calling for: a) The selection of a treatment group (Title I pupils) and a control group (regular classroom pupils) in light of the pretest needs assessment. Any pupil whose performance is higher than a preset cut-off point is assigned to the regular instructional program only; any pupil whose performance is lower than the criterion is eligible for remedial Title I instruction as a supplement to the regular curriculum; b) The assessment of the impact of the regular classroom instruction alone, in view of control group pupils' performance. By projection an average score on the posttest is estimated for the treatment group (Title I students). This average is an index of what Title I pupils achievement would be like without the supplemental instruction; c) The comparison of the predicted or expected score to the actual mean score achieved by the treatment group on the posttest. The difference is analyzed for statistical significance with a t-test.

It takes a simple conceptual step to extend that model, commonly used for a one-year/one-stage evaluation, to deal with a two-year/two-stage intervention. Two adjustments are necessary: a) In terms of the research design, one just has to redefine the treatment group and the control group. The treatment group is to include students who have been Title I participants for two consecutive years (1979-80 and 1980-81), while the control group is made of pupils who have not been in a Title I (or SCE) program during the same period; b) Thus, the determination of the expected posttest score for 1981 becomes a function of both the 1979 and 1980 achievement scores. More importantly, interpretation of the results rests not on a straightforward regression but on a path analytic approach. (A brief discussion of path analysis is offered in Appendix A). The reason for preferring the latter approach is simple: student achievement in 1979 influences their 1981 achievement not only directly but also indirectly through their 1980 performance. To give an example, one would say that, if a pupil had difficulty learning consonant clusters in the 2nd grade, that problem may not only slow his acquisition of word recognition skills in the 3rd grade, but also later, in the fourth grade, his reading speed (direct effect) as well as his reading comprehension skills (indirect effect through word recognition) may be impaired. That double impact of an earlier grade achievement on subsequent performance two years later - which makes good sense educationally - has to be accounted for in the statistical procedure adopted for evaluation.

The analysis bears on the scores of pupils presently in grade 4, 5, and 6. The results are presented in 7 tables. Table B1 shows the breakdown of the various correlation coefficients into direct and indirect causal indexes, for all three grades. Tables B2 to B4 report the following summary statistics for each grade level: the pretest (1979), mid-test (1980) and posttest (1981) means, with their respective standard deviations; the anticipated posttest mean for the treatment group; the mean "gain" or average pretest-to-posttest score change in both standard scores and normal curve equivalents (NCE's).

PRESENT GRADE	SKILL	YEAR	CORR. WITH 1980	CORR. WITH 1981	DIRECT CAUSAL EFFECT 1981		INDIRECT CAUSAL EFFECT 1981	JOINT ASSOCIAT. EFFECT (81)	TOTAL CAUSAL EFFECT (81)
					BETA	b			
4	Reading	1979	.383	.496	.358	.397	.138	---	.496/.55
		1980	1.00	.499	.362	.360	---	.137	.362
	Math	1979	.542	.573	.330	.316	.243	---	.573/.55
		1980	1.00	.627	.450	.368	---	.179	.450
5	Reading	1979	.463	.459	.228	.269	.230	---	.458/.57
		1980	1.00	.604	.498	.573	---	.105	.498
	Math	1979	.605	.652	.366	.303	.286	---	.652/.54
		1980	1.00	.693	.474	.454	---	.223	.474
6	Reading	1979	.571	.611	.407	.418	.203	---	.610/.62
		1980	1.00	.589	.356	.311	---	.232	.356
	Math	1979	.660	.662	.426	.450	.234	---	.660/.70
		1980	1.00	.638	.356	.388	.---	.282	.356

TABLE B1 - Path Analytic Data for Analysis of Cumulative Effects of Title I Program in Reading and Math

Results

A - Reading

Grade 4

To evaluate the cumulative impact of the Language Experience Program in grade 4, the MAT reading scores were examined for a total of 1,375 pupils. Of those, 585 were Title I participants in both the 1979-80 and 1980-81 school years; the remaining 790 were not enrolled in any remedial program during that same period. The latter group served as a control group.

It was found that the Title I participants had an average reading score of 46.70 in May 1979, compared to a mean of 59.34 for the regular classroom students. This represents an initial gap of 12.64 between the two subgroups. By May 1980, the mean score for the children in the treatment subgroup was up to 55.74, compared to an average of 68.14 for those in the comparison group. Analysis of the set of scores obtained by the latter revealed that 36% of the variance in the 1981 reading performance could be explained from the 1979 and 1980 achievement. The total impact of the 1979 scores on the achievement level of May 1981 was estimated at .55, twenty-seven percent of which (.15/55) was indirect, i.e. carried through 1980. Achievement in the last year had a .36 net causal effect on this year's. One thus needs to notice that it had less influence than the 1979 performance.

These figures helped chart out the 1981 expectations for the Title I pupils. In the absence of the supplemental instruction program during the two-year period, a difference of at least 11.4 points could be anticipated between the treatment and control groups at the end of 1981. That is to say that the reading score for Title I participants was expected to be 60.98, compared to an average of 72.4 for the control group. As it turned out, the Title I students obtained an average score 59.07 on the

final posttest (1981). Their improvement in reading was therefore less than what could be expected. The difference of -1.91 was found to be statistically significant at the .05 level, so there has been no cumulative effect from the Title I program on the reading achievement of pupils in grade 4.

Reading	Treatment Group N=585	Control Group N=790
1. Pretest mean	46.70	59.34
2. Standard deviation for (1)	5.81	8.45
3. Mid-test mean	55.74	68.14
4. Standard deviation for (3)	6.83	9.43
5. Post-test mean	59.07	72.40
6. Standard deviation for (5)	8.03	9.38
7. Expected posttest mean	60.98	
8. Difference at mean in SS	-1.91	
9. NCE value of (8)	-2.8	

TABLE B2.1 - Cummulative Effects data for pupils in Grade 4 (LEP)

Reading	Treatment Group N=611	Control Group N=925
1. Pretest mean	52.21	67.01
2. Standard deviation for (1)	5.32	9.06
3. Mid-test mean	58.66	73.92
4. Standard deviation for (3)	6.96	9.29
5. Post-test mean	62.55	79.00
6. Standard deviation for (5)	8.06	10.68
7. Expected posttest mean	62.32	
8. Difference at mean in SS	.23	
9. NCE value of (8)	.2	

TABLE B3.1 - Cummulative Effects data for pupils in Grade 5 (LEP)

Grade 5

To evaluate the cumulative impact of the Language Experience Program in grade 5, the MAT reading scores were examined for a total of 1,536 pupils. Of those, 611 were Title I participants in both the 1979-80 and 1980-81 school year; the remaining 925 were not enrolled in any remedial program during that same period. The latter group served as a control group.

It was found that the Title I participants had an average reading score of 52.21 in May 1979, compared to a mean of 67.01 for the regular classroom students. This represents an initial gap of 14.80 between the two subgroups. By May 1980, the mean score for the children in the treatment subgroup was up to 58.66, compared to an average of 73.92 for those in the comparison group. Analysis of the set of scores obtained by the latter revealed that 41% of the variance in the 1981 reading performance could be explained from the 1979 and 1980 achievement. The total impact of the 1979 scores on the achievement level of May 1981 was estimated at .539, more than fifty percent of which (.27/.53) was indirect i.e. carried through 1980. Achievement in the last year had a .57 net causal effect on this year's. One thus needs to notice that it had much more impact than the 1979 performance.

These figures helped chart out the 1981 expectations for the Title I pupils. In the absence of the supplemental instruction program during the two-year period, a difference of at least 16.68 points could be anticipated between the treatment and control groups at the end of 1981. That is to say that the reading score for Title I participants was expected to be 62.32, compared to an average of 79.0 for the control group. as it turned out, the Title I students obtained an average score of 62.55 on the final posttest (1981). Their improvement in reading was about equal to what could be expected. The difference of .13 was not statistically significant at the .05 level, so, there has been no cumulative effect from the Title I program on the reading achievement of pupils in grade 5.

Grade 6

To evaluate the cumulative impact of the Language Experience Program in grade 6, the MAT reading scores were examined for a total of 1403 pupils. Of those, 575 were Title I participants in both the 1979-80 and 1980-81 school years; the remaining 828 were not enrolled in any remedial program during that same period. The latter group served as a control group.

It was found that the Title I participants had an average reading score of 55.78 in May 1979, compared to a mean of 72.19 for the regular classroom students. This represents an initial gap of 16.41 between the two subgroups. By May 1980, the mean score for the children in the treatment subgroup was up to 62.70, compared to an average of 79.88 for those in the comparison group. Analysis of the set of scores obtained by the latter revealed that 46% of the variance in the 1981 reading performance could be explained from the 1979 and 1980 achievement. The total impact of the 1979 scores on the achievement level of May 1981 was estimated at .62, thirty-two percent of which (.21/.62) was indirect, i.e. carried through 1980. Achievement in the last year had a .31 net causal effect on this year's. One thus needs to notice that it had less influence than the 1979 performance.

These figures helped chart out the 1981 expectations for the Title I pupils. In the absence of the supplemental instruction program during the two-year period, a difference of at least 15.6 points could be anticipated between the treatment and control groups at the end of 1981. That is to say that the reading score for Title I participants was expected to be 68.20, compared to an average of 83.78 for the control group. As it turned out, the Title I students obtained an average score of 69.12 on the final posttest (1981). Their improvement in reading was therefore more than what could be expected. But the difference of .92 was not statistically significant at the .05 level; so, there has been no cumulative effect from the Title I program on the reading achievement of pupils in grade 6.

Reading	Treatment Group N=575	Control Group N=828
1. Pretest mean	55.78	72.19
2. Standard deviation for (1)	6.35	9.26
3. Mid-test mean	62.70	79.88
4. Standard deviation for (3)	6.97	10.86
5. Post-test mean	69.13	83.79
6. Standard deviation for (5)	8.33	9.50
7. Expected posttest mean	68.20	
8. Difference at mean in SS	.92	
9. NCE value of (8)	1.9	

TABLE B4.1 - Cummulative Effects data for pupils in Grade 6 (LEP)

Math	Treatment Group N=403	Control Group N=790
1. Pretest mean	51.27	62.33
2. Standard deviation for (1)	6.51	8.88
3. Mid-test mean	61.73	76.85
4. Standard deviation for (3)	8.85	9.88
5. Post-test mean	71.77	84.41
6. Standard deviation for (5)	8.46	8.51
7. Expected posttest mean	72.52	
8. Difference at mean in SS	-.75	
9. NCE value of (8)	-.7	

TABLE B2.2 - Cummulative Effects data for pupils in Grade 4 (CSP)

B-Math

Grade 4

To evaluate the cumulative impact of the Computational Skills program in grade 4, the MAT math scores were examined for a total of 1,193 pupils. Of those, 403 were Title I participants in both the 1979-80 and 1980-81 school years; the remaining 790 were not enrolled in any remedial program during that same period. The latter group served as a control group.

It was found that the Title I participants had an average math score of 51.27 in May1979, compared to a mean of 62.33 for the regular classroom studentts. This represents an initial gap of 11.06 between the two subgroups. By May 1980, the mean score for the children in the treatment subgroup was up to 61.73 compared to an average of 76.85 for those in the comparison group. Analysis of the set of scores obtained by the latter revealed that 47% of the variance in the 1981 math performance could be explained from the 1979 and 1980 achievement. The total impact of the 1979 scores on the achievement level of May 1981 was estimated at .55, forty-two percent of which (.23/55) was indirect, i.e. carried through 1980. Achievement in the last year had a .39 net causal effect on this year's. One thus needs to notice that it had less influence than the 1979 performance.

These figures helped chart out the 1981 expectations for the Title I pupils. In the absence of the supplemental instruction program during the two-year period, a difference of at least 11.9 points could be anticipated between the treatment and control groups at the end of 1981. That is to say that the math score for Title I participants was expected to be 72.52, compared to an average of 84.4 for the control group. As it turned out, the Title I students obtained an average score of 71.77 on the final posttest (1981.) Their improvement in math was therefore less than what could be expected. The difference of -.75 was not statistically significant at the .05 level; so, there has been no cumulative effect from the Title I program on the math achievement of pupils in grade 4.

Grade 5

To evaluate the cumulative impact of the Computational Skills Program in grade 5, the MAT math scores were examined for a total of 1,432 pupils. Of those, 611 were Title I participants in both the 1979-80 and 1980-81 school years; the remaining 925 were not enrolled in any remedial program during that same period. The latter group served as a control group.

It was found that the Title I participants had an average math score of 58.18 in May 1979, compared to a mean of 73.38 for the regular classroom students. This represents an initial gap of 15.20 between the two subgroups. By May 1980, the mean score for the children in the treatment subgroup was up to 71.53, compared to an average of 86.64 for those in the comparison group. Analysis of the set of scores obtained by the latter revealed that 56% of the variance in the 1981 math performance could be explained from the 1979 and 1980 achievement. The total impact of the 1979 scores on the achievement level of May 1981 was estimated at .54, forty-four percent of which (.23/54) was indirect, i.e. carried through 1980. Achievement in the last year had a .45 net causal effect on this year's. One thus needs to notice that it had less influence than the 1979 performance.

These figures helped chart out the1981 expectations for the Title I pupils. In the absence of the supplemental instruction program during the two-year period, a difference of at least 15.0 points could be anticipated between the treatment and control groups at the end of 1981. That is to say that the math score for Title I participants was expected to be 75.35, compared to an average of 90.4 for the control group. As it turned out, the Title I students obtained an average score of 76.42 on the final posttest (1981). Their improvement in math was therefore more than what could be expected. The difference of 1.07 was found to be statistically significant at the .05 level, so, there has been a cumulative effect from the Title I program on the math achievement of pupils in grade 5.

Math	Treatment Group N=611	Control Group N=925
1. Pretest mean	58.18	73.38
2. Standard deviation for (1)	6.62	9.89
3. Mid-test mean	71.54	86.64
4. Standard deviation for (3)	7.42	8.49
5. Post-test mean	76.43	90.40
6. Standard deviation for (5)	7.90	8.19
7. Expected posttest mean	75.35	
8. Difference at mean in SS	1.07	
9. NCE value of (8)	4.6	

TABLE B3.2 - Cummulative Effects data for pupils in Grade 5 (CSP)

Math	Treatment Group N=473	Control Group N=828
1. Pretest mean	68.91	84.74
2. Standard deviation for (1)	6.56	8.68
3. Mid-test mean	77.04	91.71
4. Standard deviation for (3)	7.55	8.41
5. Post-test mean	83.62	96.16
6. Standard deviation for (5)	6.19	9.15
7. Expected posttest mean	79.44	
8. Difference at mean in SS	4.18	
9. NCE value of (8)	7.3	

TABLE B4.2 - Cummulative Effects data for pupils in Grade 6 (CSP)

Grade 6

To evaluate the cumulative impact of the Computational Skills Program in grade 6, the MAT math scores were examined for a total of 1,301 pupils. Of those, 473 were Title I participants in both the 1979-80 and 1980-81 school years; the remaining 828 were not enrolled in any remedial program during that same period. The latter group served as a control group.

It was found that the Title I participants had an average math score of 68.90 in May 1979, compared to a mean of 84.73 for the regular classroom students. This represents an initial gap of 15.83 between the two subgroups. By May 1980, the mean score for the children in the treatment subgroup was up to 77.04, compared to an average of 91.71 for those in the comparison group. Analysis of the set of scores obtained by the latter revealed that 51% of the variance in the 1981 math performance could be explained from the 1979 and 1980 achievement. The total impact of the 1979 scores on the achievement level of May 1981 was estimated at .70, thirty-five percent of which (.25/.70) was indirect, i.e. carried through 1980. Achievement in the last year had a .39 net causal effect on this year's. One thus needs to notice that it had less influence than the 1979 performance.

These figures helped chart out the1981 expectations for the Title I pupils. In the absence of the supplemental instruction program during the two-year period, a difference of at least 16.7 points could be anticipated between the treatment and control groups at the end of 1981. That is to say that the math score for Title I participants was expected to be 79.44, compared to an average of 96.2 for the control group. As it turned out, the Title I students obtained an average score of 83.62 on the final posttest (1981). Their improvement in math was therefore more than what could be expected. The difference of 4.18 was found to be statistically significant at the .05 level, so, there has been a cumulative effect from the Title I program on the math achievement of pupils in grade 6.

Treatment Discontinuity: Carry-over Effects

In any given year, some of the pupils enrolled in a remedial instruction program such as Title I, succeed in wiping out (some of) their deficiency, to the point that they seem no longer in need of the additional support. Their performance at the end of the intervention period clearly places them above the cut-off mark set for placement into a special program during the following year. To illustrate, if a Title I pupil obtains a MAT reading score of 65 at the end of grade 4 and the criterion for Title I placement in grade 5 is a mark of 62, this pupil will not be assigned again to the special program in the following year. Such progress may be taken as evidence of the project's effectiveness. But the evidence is only partial if it is not known how much of the gain realized is maintained in the next higher grade. The question is: once the supplemental instruction is discontinued, does student reading or math achievement fall back to the level it was at before program participation?

DESIGN

To determine whether the Title I program had achieved that objective, a norm-referenced model, known as Evaluation Model-A, was adopted. The rationale underlying the model is simple: if no supplemental instruction were provided to the students in need, their status from year-to-year would be exactly the same. For instance, pupils who obtain a reading percentile rank of 15 in grade 3 are likely to score at the 15th percentile again in grade 4; if there is any change in that standing, it is attributed to the impact of the remedial instruction program.

One of the technical requirements for application of this model is that the same instrument should not be used for program placement and pretesting; (this is necessary to avoid the regression effect on the post-test scores). Since, in the practical setting, the evaluator is often faced with that problem, Baker (1981) proposed a corrective formula that would tend to minimize the regression effect. "The procedure estimates the average score for the Title I students that would have been found if a separate

pretest had been administered." (Baker, 1981, p. 1). It is that "expected pretest score" that forms the baseline for subsequent comparison(s). To make the results of the comparison immediately interpretable, all scores need to be converted into NCE's. In extending the norm-referenced model for studying carry-over effects, one has to focus on a group of pupils who had been Title I participants in 1979-90, but were not in 1980-81. The program's immediate impact can be assessed for the first half of that period, by comparing the average 1980 performance to the 1979 average achievement score. The procedure is then repeated for the 1979 and 1981 mean scores (in NCE). If there is no carry-over effect from the first year (of remedial support) to the second year (no remedial support), the two scores should be comparable. But if the latter score is significantly higher than the former, it is inferred that the difference is a residual of the gain achieved through program participation. The analysis bears on the scores of pupils presently in grade 4, 5, and 6. The results are reported in Table B5.

SKILL	GRADE	MEAN NCE SCORE '79	MEAN NCE SCORE '80	MEAN NCE SCORE '81	DIFF. 79-80	DIFF. 80-81	DIFF. 79-81
Reading	4	41.3	50.0	39.0	8.7	-11.0	-2.3
	5	39.0	42.5	32.3	3.5	-10.2	-6.7
	6	32.2	36.5	33.7	4.2	-2.8	1.4
Math							
	4	41.3	48.9	41.3	7.6	-7.6	0
	5	32.3	43.6	40.1	11.3	-3.5	7.8
	6	32.3	42.5	37.7	10.2	-4.8	5.4

TABLE B5 - Mean Scores and progress on MAT of Title I pupils for (3) consecutive years - Carry-over effects are indicated in last column

Results

A-Reading

Grade 4

To evaluate the carry-over effects of the Title I Language Experience Program in grade 4, the MAT test data were examined for a group of 161 children who had been Title I participants in the year 1979-80, but not in 1980-81. They had an average reading score of 48.87 at the end of 1979, approximately 5 points lower than that of the total group of children of the same grade in the Newark public schools. After proper adjustment for pretesting effects, their performance could be established at the NCE level of 41.3. At the conclusion of a successful year in the Title I program (1980), their mean score moved up to 61.11. That mark was more than 3 points higher than the criterion of 58 set for placement of a fourth grader in the Title I remedial program. It corresponded to a NCE score of 50.0, yielding a gain of almost 9 points.

In the ensuing year, however, when the remedial support was discontinued, the same rate of progress could not be maintained by these pupils. Their 1981 mean reading score is 63.34, which translates into a NCE of 39.0. Not only had there been no gain over the 1980 performance, but the pupils' standing fell below their 1979 level by -2.3 points. The improvement in reading being less than what could be expected, one can conclude that there has been no carry-over effect from the Title I program for the children in grade 4.

Grade 5

To evaluate the carry-over effects of the Title I Language Experience Program in grade 5, the MAT test data were examined for a group of 119 children who had been Title I participants in the year 1979-80, but not in 1980-81. They had an average reading score of 54.52 at the end of 1979, approximately 6 points lower than that of the total group of children of the same grade in the Newark public schools. After proper adjustment for pretesting effects, their performance could be established at the NCE level of 39.0. At the conclusion of a successful year in the Title I program (1980), their mean score moved up to 65.53. That mark was more than 3 points higher than the criterion of 62 set for placement of a fifth grader in the Title I remedial program. It corresponded to a NCE score of 42.5, yielding a gain of almost 3.5 points.

In the ensuing year, however, when the remedial support was discontinued, the same rate of progress could not be maintained by these pupils. Their 1981 mean reading score is 67.88, which translates into NCE of 32.3. Not only had there been no gain over the 1980 performance, but the pupils' standing fell below their 1979 level by-6.7 points. The improvement in reading being less than what could be expected, one can conclude that there has been no carry-over effect from the Title I program for the children in grade 5.

Grade 6

To evaluate the carry-over effects of the Title I Language Experience Program in grade 6, the MAT test data were examined for a group of 87 children who had been Title I participants in the year 1979-80, but not in 1980-81. They had an average reading score of 58.77 at the end of 1979, approximately 6 points lower than that of the total group of children of the same grade in the Newark public schools. After proper adjustment for pretesting effects, their performance could be established at the NCE level of 32.3. At the conclusion of a successful year in the Title I program (1980), their mean score moved up to 69.88. That mark was more than 2 points higher than the criterion of 67 set for placement of a sixth grader in the Title I remedial program. It corresponded to a NCE score of 36.5 yielding a gain of almost 4 points.

In the ensuing year, however, when the remedial support was discontinued, the same rate of progress could not be maintained by these pupils. Their 1981 mean reading score is 74.31, which translates into NCE of 33.7. There has been no gain over the 1980 performance, but the pupils' standing is above their 1979 level by 1.4 points. The improvement in reading being more than what could be expected, one can conclude that there has been a carry-over effect from the Title I program for the children in grade 6.

B-Math

Grade 4

To evaluate the carry-over effects of the Title I Computational Skill Program in grade 4, the MAT test data were examined for a group of 95 children who had been Title I participants in the year 1979-80, but not in 1980-81. They had an average math score of 54.18 at the end of 1979, approximately 3 points lower than that of the total group of children of the same grade in the Newark public schools. After proper adjustment for pretesting effects, their performance could be established at the NCE level of 41.3. At the conclusion of a successful year in the Title I program (1980), their mean score moved up to 70.38. That mark was more than 5 points higher than the criterion of 65 set for placement of a fourth grader in the Title I remedial program. It corresponded to a NCE score of 48.9 yielding a gain of almost 8 points.

In the ensuing year, however, when the remedial support was discontinued, the same rate of progress could not be maintained by these pupils. Their 1981 mean math score is 77.54, which translates into NCE of 41.3. Not only had there been no gain over the 1980 performance, but the pupils' standing is equal to their 1979 level. The improvement in math being no higher than what could be expected, one can conclude that there has been no carry-over effect from the Title I program for the children in grade 4.

Grade 5

To evaluate the carry-over effects of the Title I Computational Skills Program in grade 5, the MAT test data were examined for a group of 85 children who had been Title I participants in the year 1979-80, but not in 1980-81. They had an average math score of 78.94 at the end of 1979, approximately 6 points lower than that of the total group of children of the same grade in the Newark public schools. After proper adjustment for pretesting effects, their performance could be established at the NCE level of 32.3. At the conclusion of a successful year in the Title I program (1980), their mean score moved up to 78.94. That mark was more than 3 points higher than the criterion of 75 set for placement of a fifth grader in the Title I remedial program. It corresponded to a NCE score of 43.6 yielding a gain of almost 11 points.

In the ensuing year, however, when the remedial support was discontinued, the same rate of progress could not be maintained by these pupils. Their 1981 mean math score is 81.63, which translates into NCE of 40.1. Although there has there been no gain over the 1980 performance, the pupils' standing went above their 1979 level, by 7.8 points. The improvement in math being more than what could be expected, one can conclude that there has been a carry-over effect from the Title I program for the children in grade 5.

Grade 6

To evaluate the carry-over effects of the Title I Computational Skills Program in grade 6, the MAT test data were examined for a group of 61 children who had been Title I participants in the year 1979-80, but not in 1980-81. They had an average math score of 71.20 at the end of 1979, more than 6 points lower than that of the total group of children of the same grade in the Newark public schools. After proper adjustment for pretesting effects, their performance could be established at the NCE level of 32.3. At the conclusion of a successful year in the Title I program (1980), their mean score moved up to 82.82. That mark was about one point higher than the criterion of 82 set for placement of a sixth grader in the Title I remedial program. It corresponded to a NCE score of 42.5 yielding a gain of almost 10 points.

In the ensuing year, however, when the remedial support was discontinued, the same rate of progress could not be maintained by these pupils. Their 1981 mean math score is 86.77, which translates into a NCE of 37.7. Though there has been no gain over the 1980 performance, the pupils' standing went above their 1979 level by 5.4 points. The improvement in math being more than what could be expected, one can conclude that there has been a carry-over effect from the Title I program for the children in grade 6.

SUMMARY & CONCLUSIONS

Regardless of the method of analysis used, or the group one focuses on, the pattern of results is the same: the long-term effects of the Title I project are more evident and consistent for pupils presently in grade 6. Indeed, for all the children at that grade level, the impact of the remedial instruction has been sustained from year to year, in both reading and math. Those who participated for only one year have been able to maintain at least one-third of their initial gain; those in need of longer assistance (two-year enrollment) have also shown a strengthening of their skills.

In grade 5, sustaining effects have been detected for the Computational Skills program, but not for the Language Experience program. In reading, while the pupils who enrolled for two years have simply not performed beyond the expected level, children in the other group, despite a promising start in 1979-80, seem to have fallen below their initial standing, once the remedial support was discontinued. The math performance is clearly stronger. Both subgroups have significantly improved over their third-grade standing, and the carry-over effect (the largest observed) is no less than 69% of the gain achieved the previous year. In grade 4, the Title I program has shown no sustaining effects in either reading or computational skills. Even after two years of remedial instruction, these pupils still score below the level of minimal expectation. As for those with only one year of Title I experience, the scores they obtain now do not reflect their initial gains; they are no higher than what could have been realized through the regular classroom instruction alone. The findings carry three major implications:

1. Comparison of the change in math achievement to that in reading achievement suggests that the Computational Skills program may be in the long-run more effective than the Language Experience program.

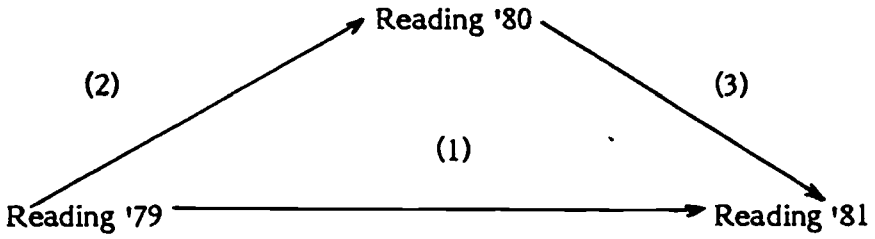
2. The overall pattern of results obtained in the analysis of sustaining effects runs contrary to that commonly observed in the evaluation of short-term effects. Usually, the immediate impact of the remedial project is found to be stronger in the early years and to decline as one moves up the grades. When it comes to cumulative or carry-over

effects, the trend seems to be in the opposite direction. One may try to link this finding to variations in the structure of the curriculum: the nature or organization of the curriculum may change more significantly from the second to the fourth grade than it does from the fourth to the sixth grade; and as a result, it becomes more difficult for children in the early grades to maintain their gains from year to year. A task is thus defined for curriculum specialists and researchers to establish whether such a change occurs in the structure of the curriculum between grades two and four. In the meantime, it is recommended that the project's administrators devote great attention to the fourth grade level, in an effort to identify factors responsible for the consistent underachievement.

3. Finally, in regard to evaluation methodology, it would be beneficial if program evaluators broaden their scope on student previous achievement to include not just the most recent performance. The net causal effect of achievement two years earlier on present-day reading or math scores ranges from .27 to .45. When that direct impact is combined with the indirect one transmitted from one year to the next, the significance of earlier achievement often exceed that of the most recent performance. Evaluators, program operators and funding sources need to keep that in view when they are defining expectations about project outcomes.

APPENDIX A

Note 1: Path Analysis. Path analysis is a statistical procedure which is usually applied to test the logical implications of a theoretical model. The model is to define the relationships and (at least a weak) causal order among variables in a set. The technique allows for the decomposition of the covariation between any pair of variables into a causal and a non-causal component. The causal component may be broken down into a direct effect index and one or more indirect effect indexes. The non-causal component may also be bro' en down into a spurious-effect index and a joint-association index. This detailed analysis of each correlation coefficient provides a great deal more information than what could be obtained through a straight (multiple) regression approach. The diagram below illustrates the use of the technique in the present study:



(1) = direct impact of 1979 performance on 1981 performance (3) = direct impact of 1980 performance on 1981 performance (2) x (3) = indirect impact of 1979 performance on 1981 performance Application of the technique, as a tool for a Model-C type of evaluation, affects mainly the calculation of the expected posttest score for the treatment group. In the regular Model-C1 analysis, the formula for the expected score is:

$$\overline{Y}'_t = \overline{Y}_c + b_c (\overline{X}_t - \overline{X}_c)$$

In the extended model, that we may call Model-C3, a new component is introduced into the formula, to represent the total impact of the 1979 performance. The expected postest mean is calculated as:

$$\overline{Y'}_t = \overline{Y}_c + b_{3c}(\overline{X}_{2t} - \overline{X}_{2c}) + (b_{1c} + b_{23c})(\overline{X}_{1t} - \overline{X}_{1c})$$

where $\overline{Y'}_t$ = predicted posttest mean for treatment group (81)

$\overline{Y}_c$ = posttest mean for control group (81)

$\overline{X}_2$ = mean test scores in 1980

$\overline{X}_1$ = mean test scores in 1979

b_{3c} = direct causal effect of 1980 on 1981 scores

b_{1c} = direct causal effect of 1979 on 1981 scores

b_{23c} = indirect causal effect of 1979 on 1981 scores

c = subscript indicating 'control group'

t = subscript indicating 'treatment group'

The total impact of the 1979 achievement level, which combines its direct and its indirect causal components, is first calculated using the standardized coefficients. As can be seen in Table B1, it is equivalent to the zero-order correlation. To transform that value into an unstandardized regression coefficient (needed in the above formula), one simply has to remember that:

$$b_{yx} = \frac{r_{xy}\, s_y}{s_x}$$

Note 2: Adjusted Model - A

The following formula is used to adjust the mean pretest score for a Model-A analysis, when a separate set of scores is not available for program selection.

$$\overline{X}_{p'} = \overline{X}_t - r_{xx}(\overline{X}_t - \overline{X}_p) \text{ where}$$

$\overline{X}_{p'}$ = the expected mean score of the Title I group on the pretest if a separate pretest had been administered;

$\overline{X}_p$ = the mean score of the Title I group on the selection pretest;

$\overline{X}_t$ = the mean score of the total group (from which the Title I students were selected) on the pretest selection. The determination of the number of Title I students selected per grade was based on the number of participants in the Title I project in that grade. Individual members in the non-Title I group were then selected by using a table of random numbers; and

r_{xx} = the test/re-test reliability for the total group.

APPENDIX B

SCHOOL _____ PROJECT COORDINATOR _____ PRINCIPAL _____

PROJECT SETTING DESCRIPTION

DESCRIPTOR / GRADE		PRE K /K	1	2	3	4	5	6	7	8	9
	RC										
READING	PO										
	LAB										
	RC										
COMPUTATION	PO										
	LAB										
	RC										
P.E.P	PO										
	LAB										
	RC										
OTHER*	PO										
	LAB										

*Specify: _____ (This category is for any other project going on at the site, that aims at the cognitive and/or non-cognitive development of Title I participants. The Pre K/K project may be classified in that category.)

RC - Classroom
PO - Pull-out
LAB - Laboratory

References

Baker, R. F. Practical Guide to Model-A Statistical Adjustment. Washington, D.C.: NTS, February 1981

Bronfenbrenner, U. A Report on Longitudinal Problems: Is Early Education Effective? Department of HEW, Washington, D.C. 1974

Festinger, L. Theory of Social Processes. Human Relations, 7, 117-140, 1954

Glass, G.V. & Smith, M.L. "Pull-out" in Compensatory Education. U.S. Office of Education, Washington, D.C. 1979

Kirk, R.E. Experimental Design: Procedures for the Behavioral Sciences. Brook/Cole: Belmont, California 1968

Merton, R.K. Social Theory and Social Structure. Glencoe: Free Press, 1957

National Instute of Education. Evaluation of Compensatory Education: An Interim Report. Washington, D.C. 1976

National Institute of Education. Compensatory Education Services Washington, D.C. 1977

Tallmadge, K. & Horst, D.P. Practical Guide to Measuring Project Impact on Student Achievement. Department of HEW, Washington, D.C. 1975

Tallmadge, K. & Wood, C.T. User's Guide: ESEA Title I Evaluation and Reporting System. Department of HEW, Washington, D.C. 1978.